databricks

# Get Started with Databricks for Business Leaders

# Agenda





An introduction to Databricks

Working with Databricks teams

How Databricks supports governance and security

How Databricks brings down the cost of ownership

Next steps

# An introduction to Databricks

# Data is everywhere.

# A successful data and AI strategy

Processes

People

Platform

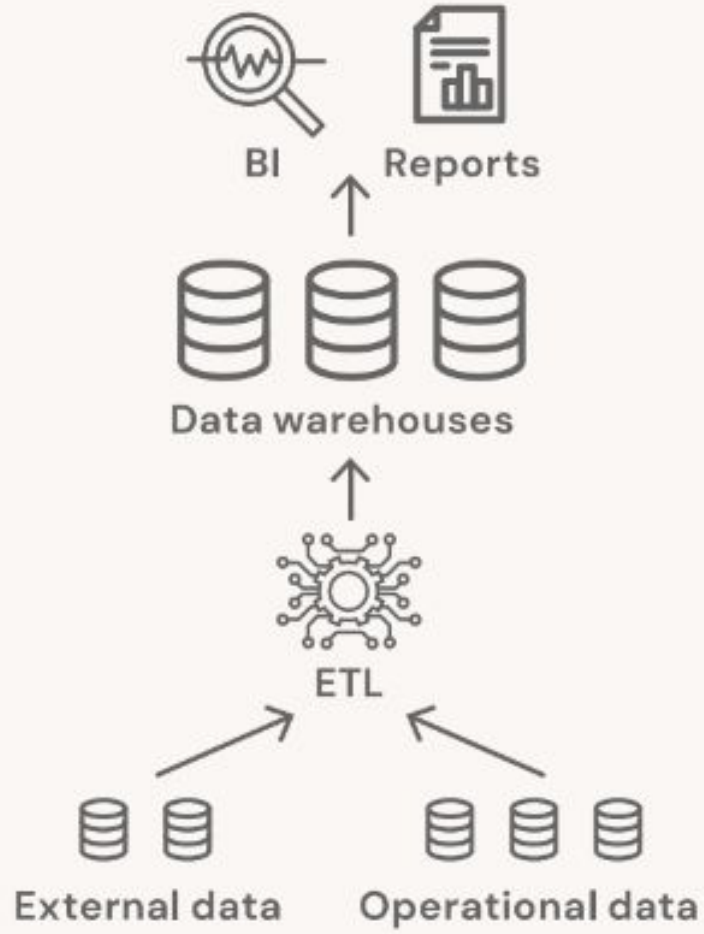

databricks


APACHE
Spark™


DELTA LAKE


mlflow™

# Data warehouse




Pros:

- Business intelligence (BI)
- Analytics
- Structured & clean data
- Predefined schemas

Cons:

- No support for semi or unstructured data
- Inflexible schemas
- Struggles with volume and velocity upticks
- Long processing time

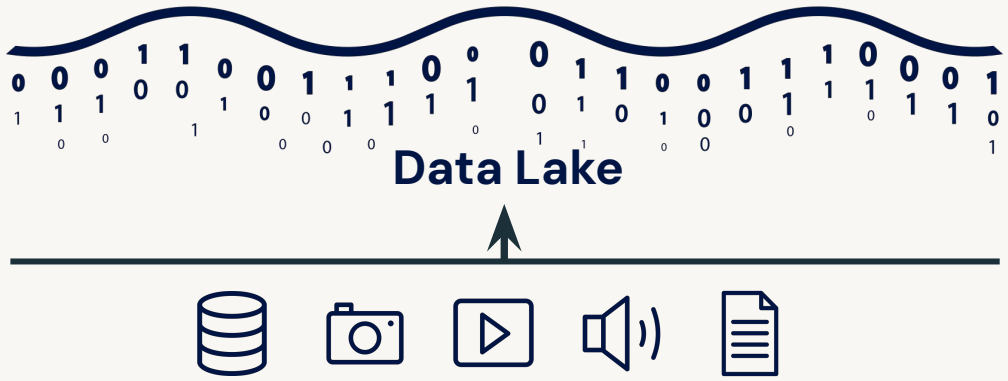
# Data Lakes

Pros:

- Flexible data storage
- Streaming support
- Cost efficient in the cloud
- Support for AI and Machine Learning

Cons:

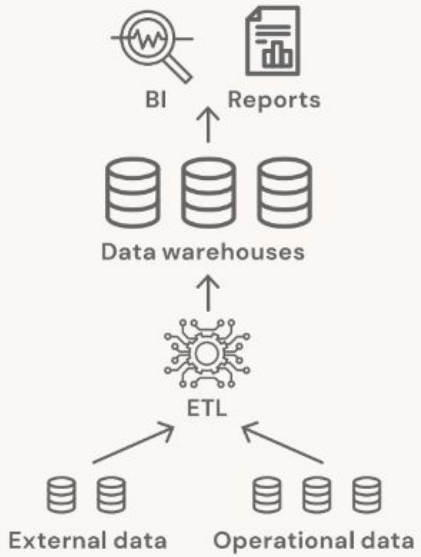
- No transactional support
- Poor data reliability
- Slow analysis performance
- Data governance concerns
- Data warehouses still needed



# Data lakehouse

One platform to unify all your data, analytics and AI workloads

## Data warehouse



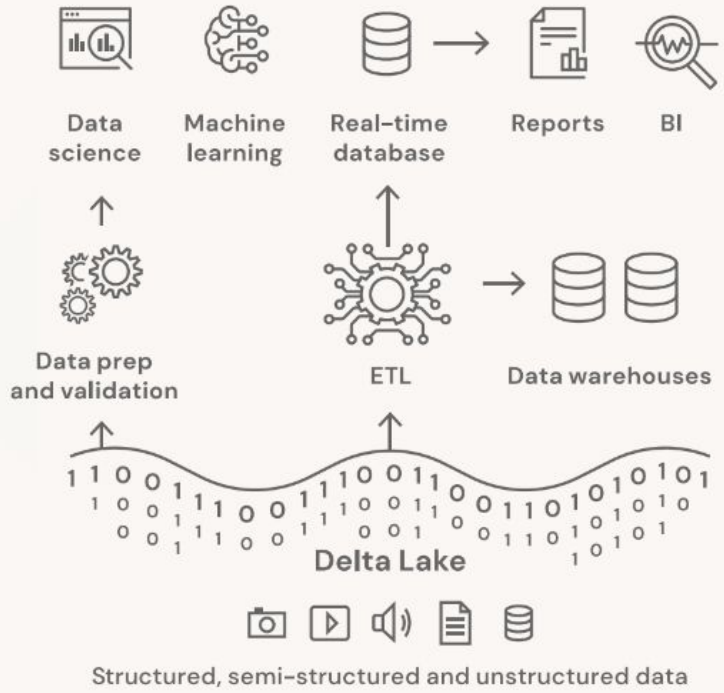

## Lakehouse Platform

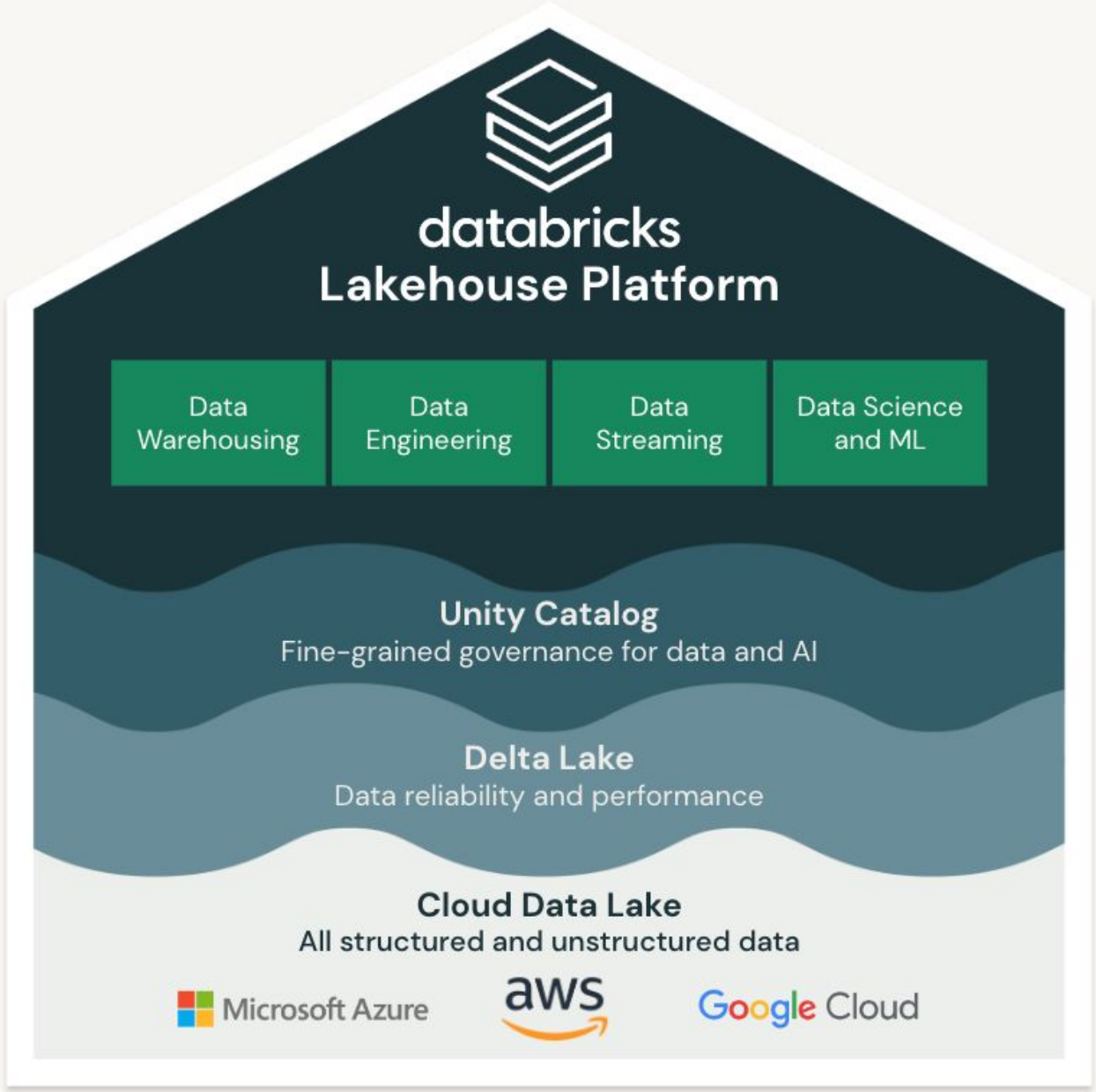
All machine learning, SQL, BI, and streaming use cases

One security and governance approach for all data assets on all clouds

An open and reliable data platform to efficiently handle all data types

## Delta Lake






# Databricks Lakehouse Platform

## Simple

Unify your data warehousing and AI use cases on a single platform

## Open

Built on open source and open standards

## Multicloud

One consistent data platform across clouds

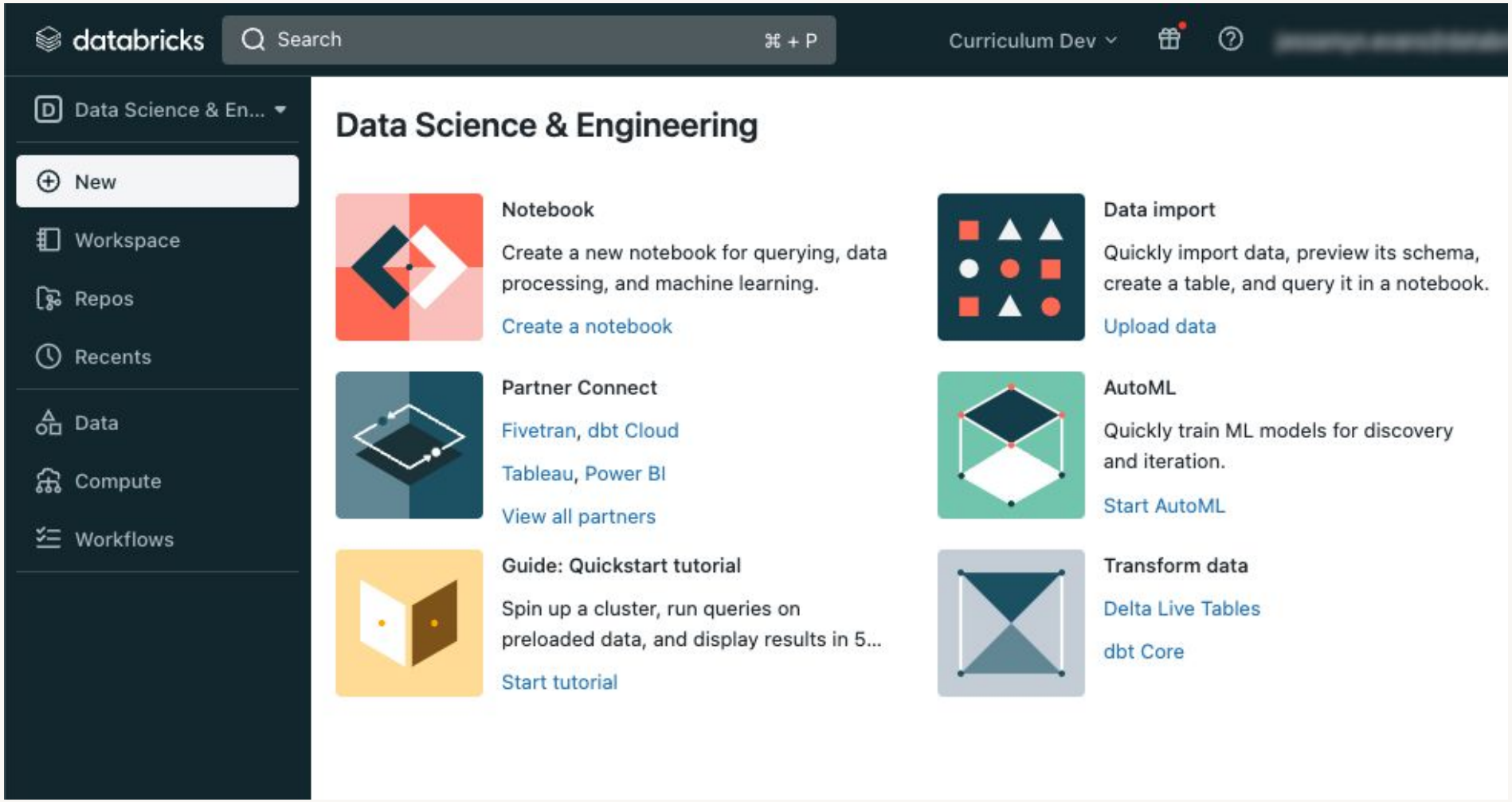
# Persona-based services

Data Analyst

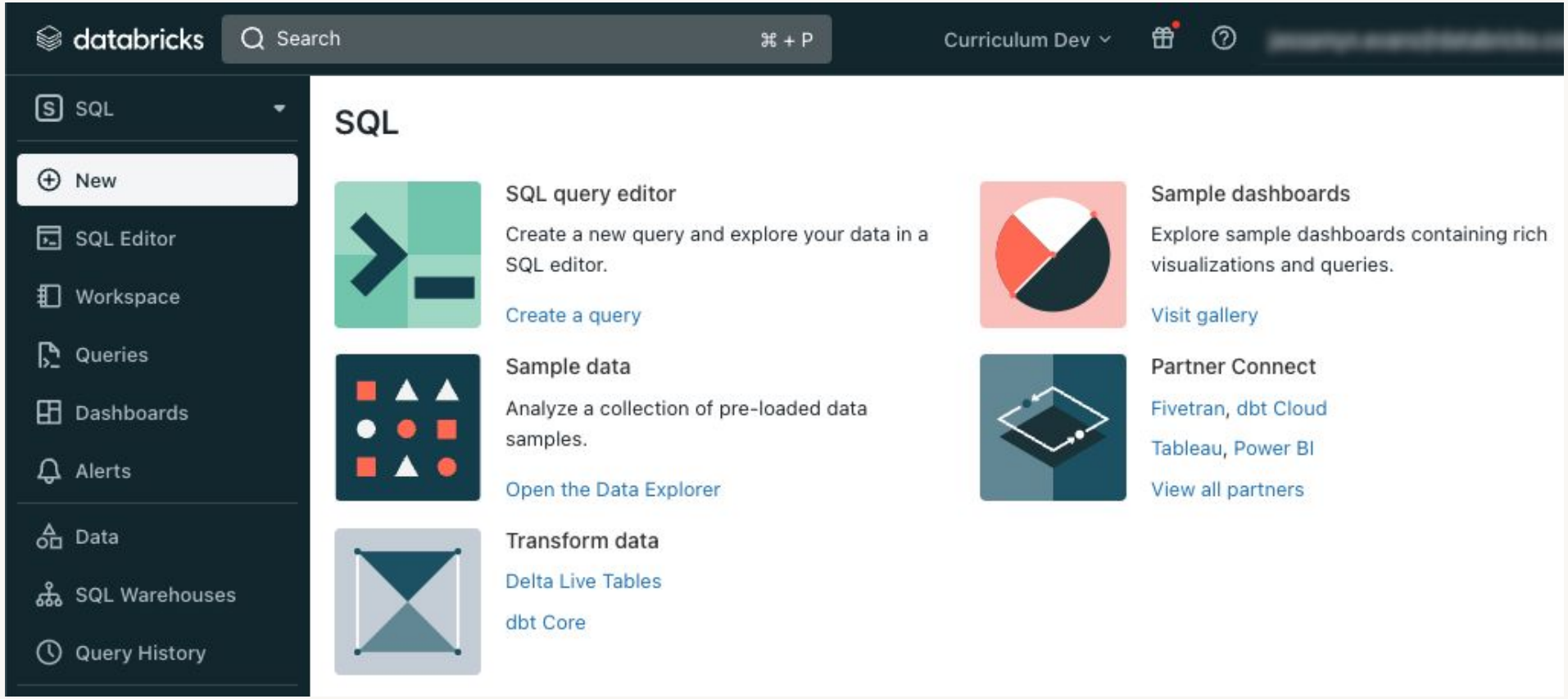
Data Engineer

Data Scientist

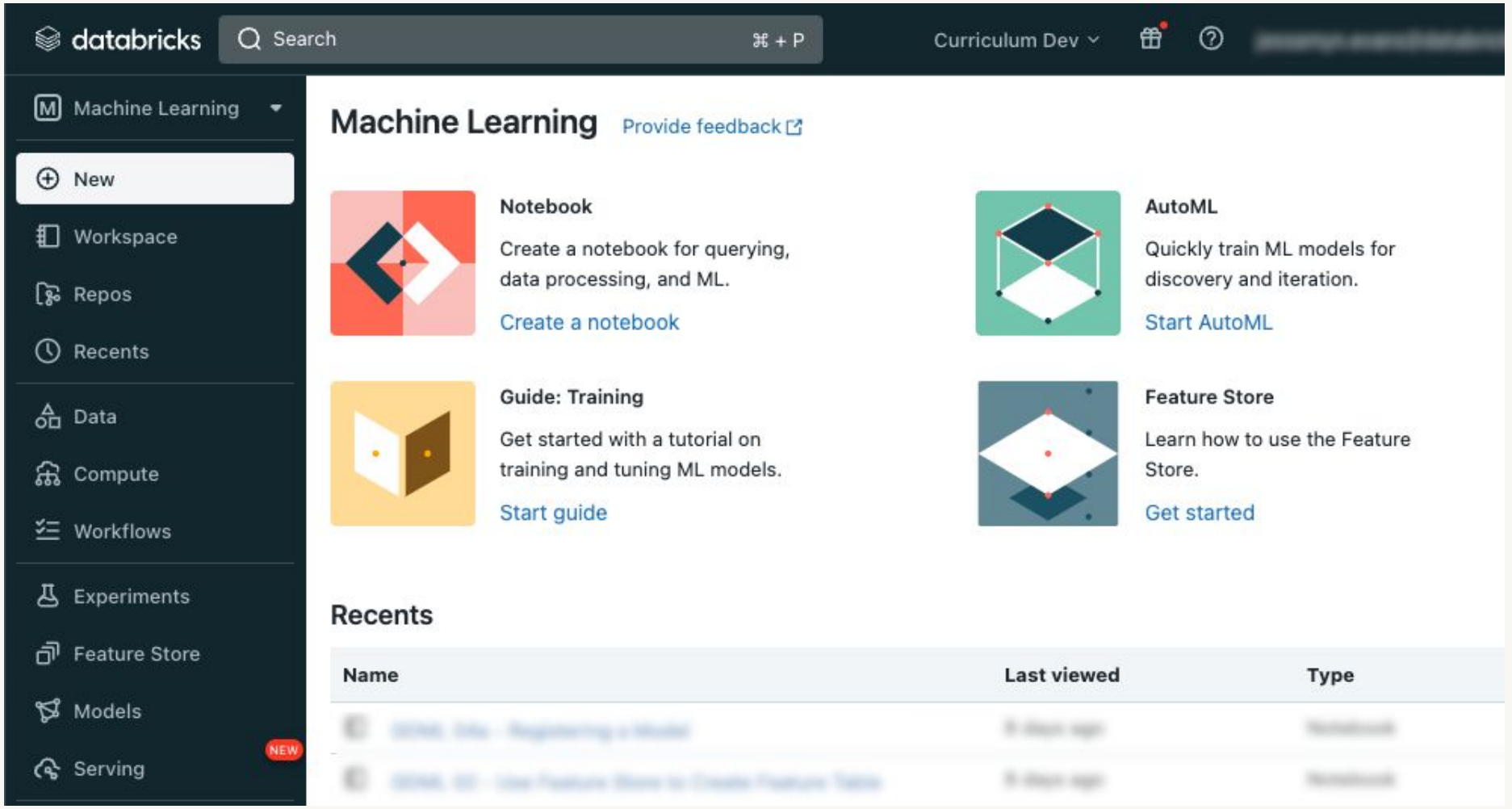
Data Engineer



Data Analyst



Data Scientist



# Support the workloads of your data team

## To create quality data, insights and predictions that drive future actions

**Data Engineering**

Ingest & transform

**Data Streaming**

Real-time insights

**Data Warehousing**

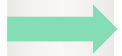
SQL & business insights

**Data Science**

ML to predict outcomes

**Data Governance**

Management and secure discovery of data assets, with auditing and lineage



Data Sharing

# Data-Forward: from siloed to unified teams

## Transforming to product-based outcomes


**Specialized team, working in silos**
Limits innovation and slows initiatives

**Unified teams, shared responsibility**
Data agility, acceleration & innovation

# The Partner Ecosystem

see more at
databricks.com/partners

## Visual ETL & Data Integration

with the Data Ingestion Network of partners with 100s of Application Sources

Azure Data Factory, Fivetran, Informatica, talend, Qlik, MATILLION, StreamSets, precisely, FISHTOWN ANALYTICS, Infoworks, WANDISCO, TRIFACTA, CONFLUENT, mongoDB.

## Business Intelligence

tableau, Power BI, Azure Synapse, Qlik, plotly, Amazon Redshift, ThoughtSpot, looker, Google BigQuery

## Machine Learning

Azure Machine Learning, Amazon SageMaker, Google AI Platform, MathWorks, RStudio, H2O.ai, dataiku

## Data Providers

dun & bradstreet, SAFEGRAPH, experian, yipitDATA, arm INSIGHT

## Centralized Governance

AWS Glue, Informatica, Alation, PRIVACERA, IMMUTA, collibra

databricks

Microsoft Azure

aws

Google Cloud

## Top Consulting & SI Partners

accenture, avanade, Insight, NEUDESIC, PARIVEDA, wipro, Infosys, Tech Mahindra, AGILISIUM, BLUEGRANITE, Hitachi Solutions, ONICA, Capgemini, Cognizant, Deloitte., ELASTACLOUD, OPENVALUE, b.telligent, GO DATA DRIVEN, kumulus, BlueShift, deep_dive, Mindtree, slalom, TATA CONSULTANCY SERVICES, JUST ANALYTICS, Intellify, Versor, ZEAL

# Customer Testimonials for the lakehouse

# coins.ph

Coins.ph uses Databricks to deliver ML-powered fraud detection and anti-money-laundering solutions while optimizing financial reconciliation.

## Use Case

- Fraud detection; anti-money laundering
- Legacy analytics system built on EMR struggled to turn massive data into meaningful insights

## Why Databricks?

- Unified platform for data teams to collaborate on data preparation, simple analytics and prototype new models
- Delta Lake ensures consistent data pipelines that feed data downstream for ML
- MLflow provides easy development and tracking of new ML models

## Impact

- **14X reduction** in complaints received
- **70% operational cost reduction** in compute costs
- **50% infrastructure cost reduction**

healthdirect Australia

Healthdirect Australia uses Databricks to improve ETL processing times and algorithm accuracies, enabling improved quality of patient care.

## Use Case

- Predictive analytics to improve the quality of patient care
- Legacy system couldn't process data efficiently, unable to provide a single source of truth across public and private sector data sources, and siloed data teams were inefficient

## Why Databricks?

- Fully managed cloud platform simplifies operations and delivers superior performance
- Delta Lake provides high data integrity and enables analysts to act on insights faster and adhere to regulatory standards
- Interactive workspaces enable data team collaboration

## Impact

- **1440x Faster ETL:** Processing a million records a minute compared to 24 hours
- Improved accuracy of fuzzy name match algorithm: **From less than 80% with manual verification to 95% and no manual intervention**

AMGEN

Amgen uses Databricks Lakehouse to accelerate the drug development lifecycle and deliver better patient outcomes.

## Use Cases

- Drug development and discovery; Supply chain optimization; Sales forecasting
- Struggled to unlock the potential of data across various organizations

## Why Databricks?

- Lakehouse unifies clinical, real-world and manufacturing data to use for ML and BI, expediting drug discovery
- Delta Lake provides ACID compliance and historical lookback, helping optimize supply chains and operations
- Databricks SQL allows live data exploration without the need for a data warehouse

## Impact

- **$10+ million** in potential savings for the business
- **280+** ML and analytics use cases
- **25% reduction** in compute costs

# Columbia

Columbia uses Databricks to democratize access to consistent and reliable data, empowering key stakeholders to make smarter business decisions.

## Use Case

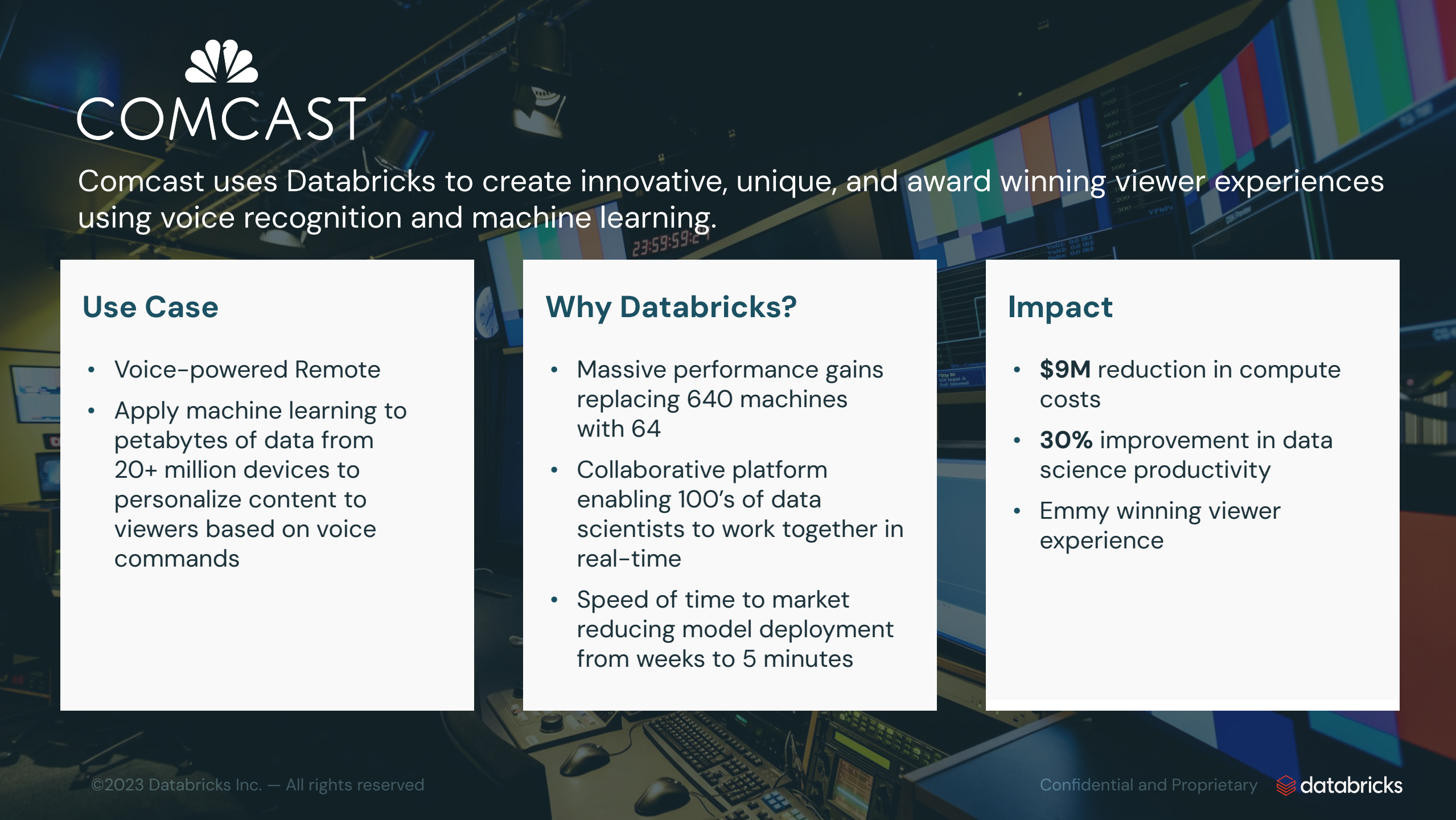
- Retail and wholesale demand forecasting
- Informatica & Teradata infrastructure siloed, lacked performance, high costs to scale (hardware, ,storage, licensing, integration)

## Why Databricks?

- Delta Lake enables all ETL and real-time architecture for developing data models
- Democratization of access to consistent, reliable data -- LOBs turn around analysis faster, less support from data engineering
- Integrated security with Azure data services: ADF, SQL DW, PBI

## Impact

- **48x faster ETL** workloads **from 4 hours to 5 min**
- **Decreased pipeline creation time by 70%**
- Reduced latency from **24-48 hours to real-time**



COMCAST

Comcast uses Databricks to create innovative, unique, and award winning viewer experiences using voice recognition and machine learning.

## Use Case

- Voice-powered Remote
- Apply machine learning to petabytes of data from 20+ million devices to personalize content to viewers based on voice commands

## Why Databricks?

- Massive performance gains replacing 640 machines with 64
- Collaborative platform enabling 100's of data scientists to work together in real-time
- Speed of time to market reducing model deployment from weeks to 5 minutes

## Impact

- **$9M** reduction in compute costs
- **30%** improvement in data science productivity
- Emmy winning viewer experience

# Working with Databricks teams

# Your Databricks support team

**Account Executive (AE)**

**Solutions Architect (SA)**

**Delivery Solutions Architect (DSA)**

# Account Executive (AE)

- Industry and region focused
- Help with developing key business use cases
- Key resource for all your Databricks questions

# Solutions Architect (SA)

- Demonstrates the platform features
- Partners with you on building out proof of value examples
- Works with the AE to realize the value of your Databricks purchase

# Delivery Solutions Architect (DSA)

- Supports adoption and consumption of Databricks
- Proactively mitigates blockers, such as security, governance, and compliance
- Facilitates your use cases going from POV to production

# Databricks Professional Services team

Migrate and connect your data

Build your lakehouse ecosystem

Innovate on your data with analytics

# Databricks provides comprehensive **Training** for the Databricks Lakehouse Platform

### Accelerate business outcomes

Achieve the full potential of your data, analytics, and AI initiatives on the lakehouse

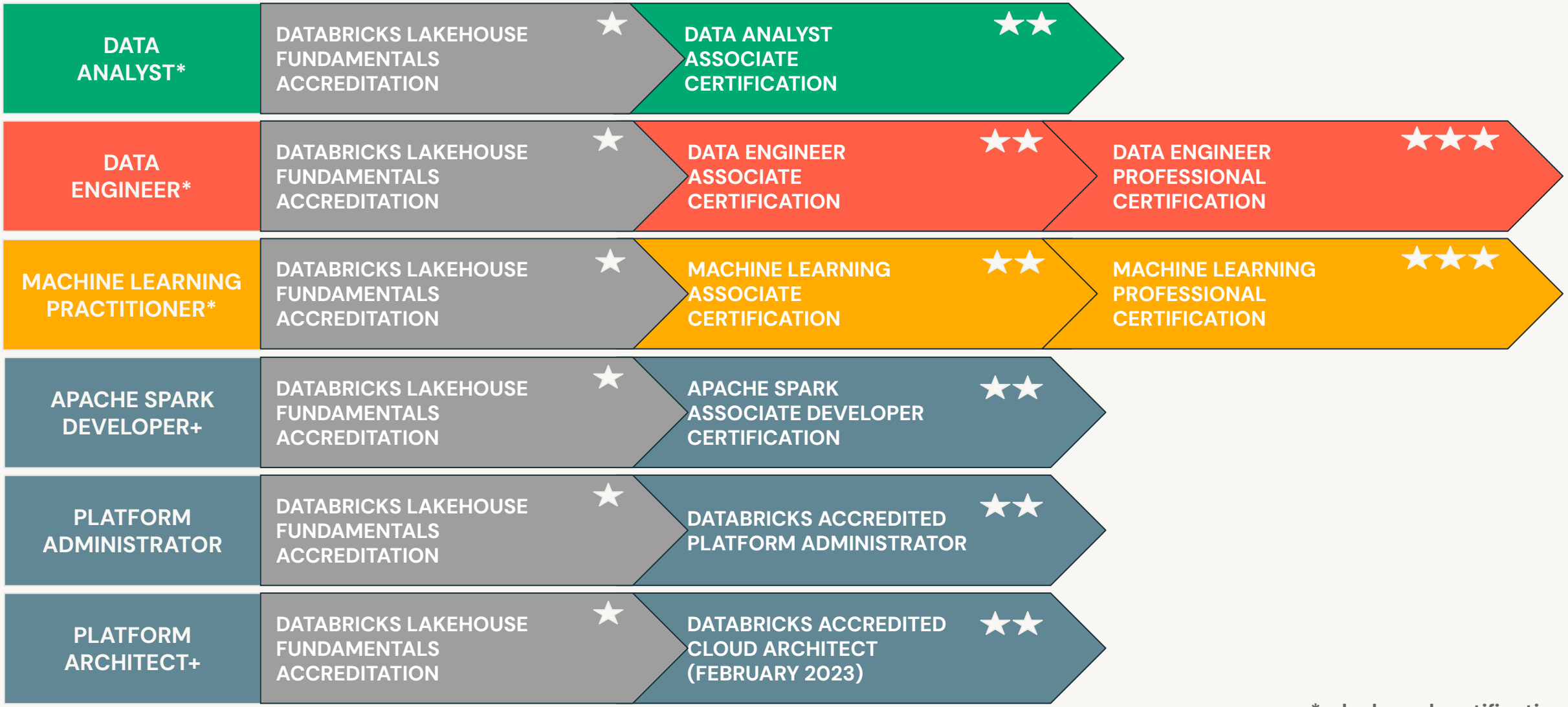
### Deploy industry-leading technology

Utilize cutting-edge and proven technology to drive results for your organization

### Foster career development

Enable your organization with role-based trainings for career acceleration and employee retention

# Databricks offers role-based learning pathways



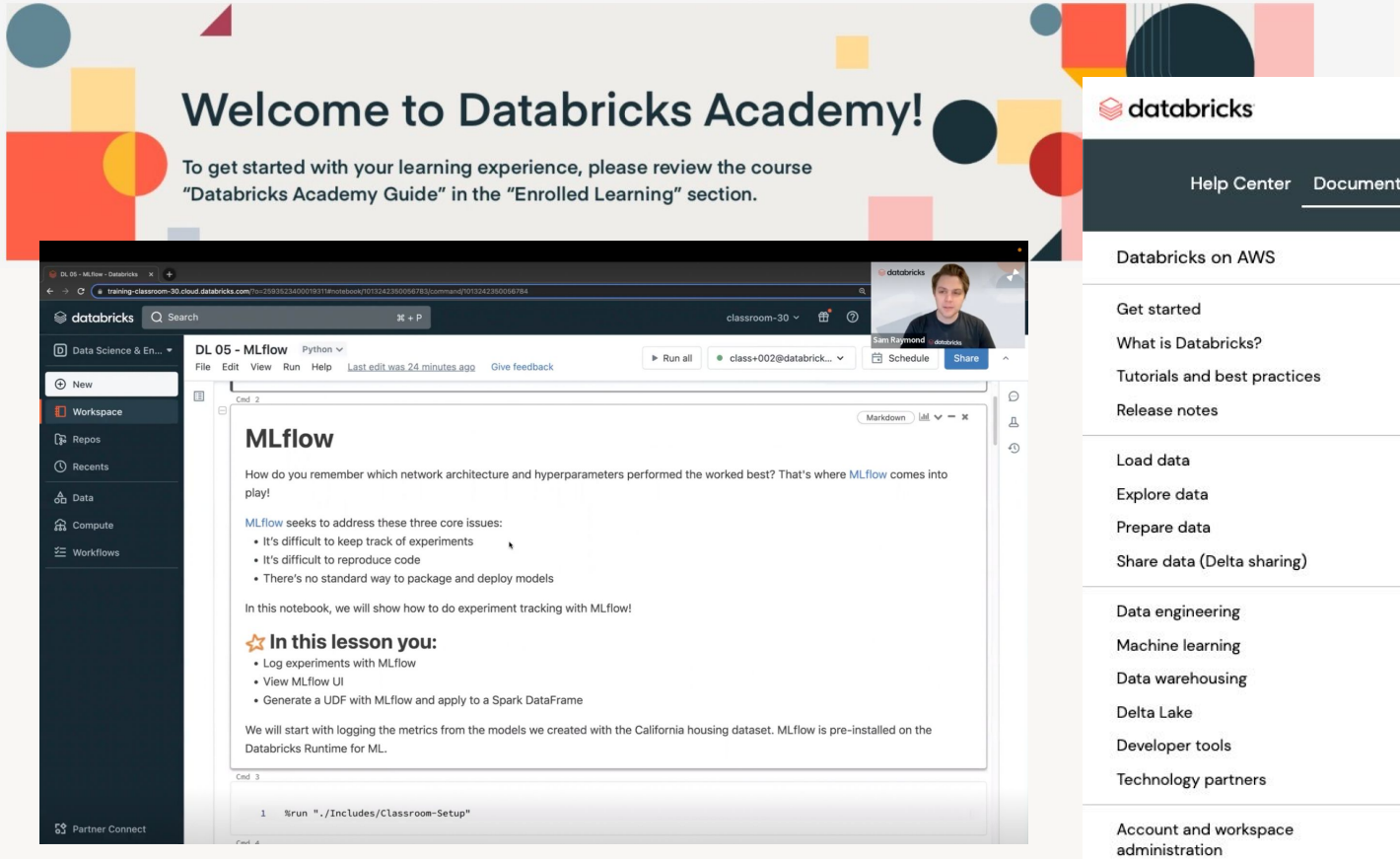


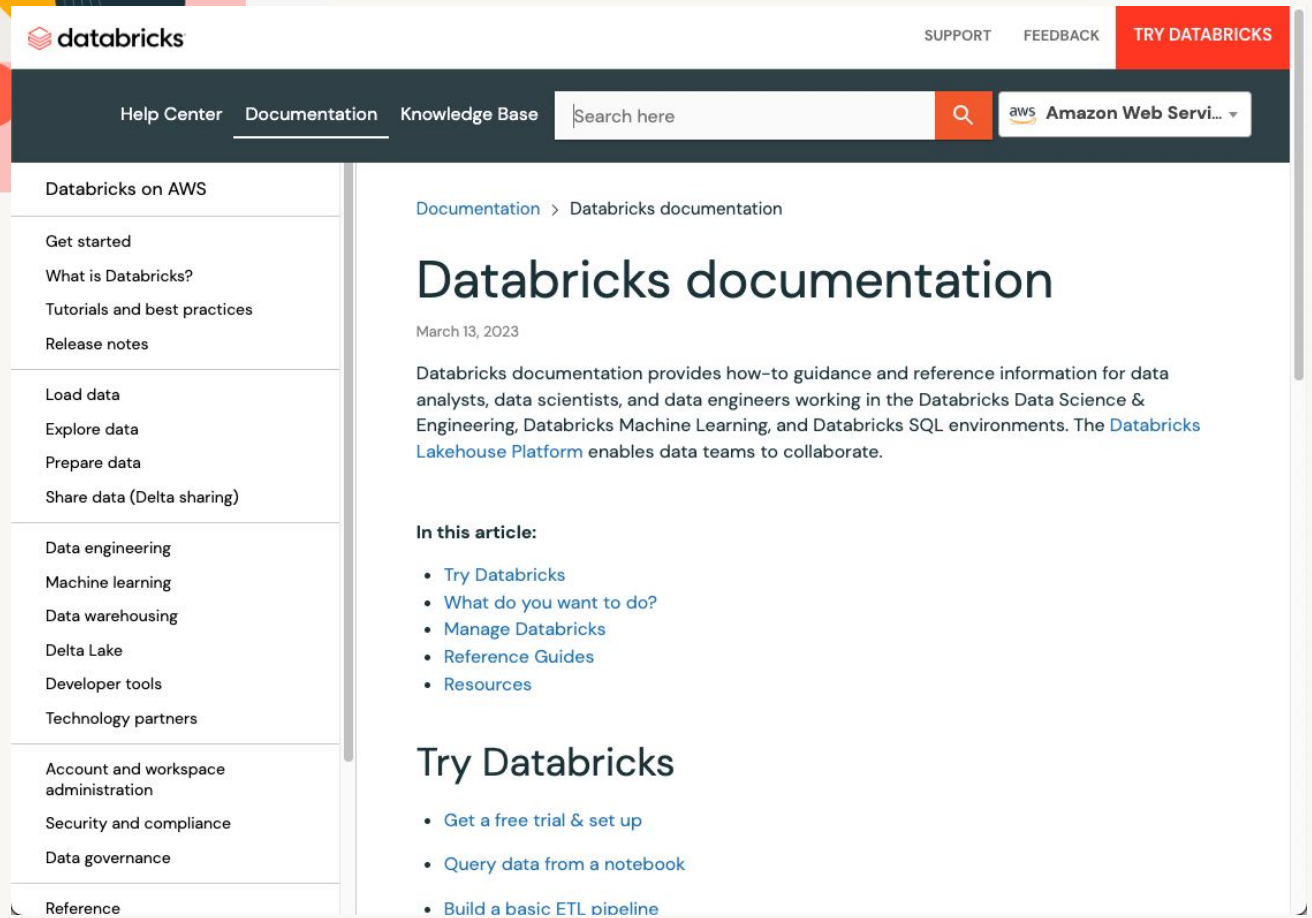



*role-based certifications
+specialty certifications

# Training and Documentation

## Databricks Academy and documentation



# How Databricks supports data governance and security

# Data governance and security

## Your data security is our priority

### Build on a secure and trusted platform

Compute, network and workload security controls with best practices guidance

### Protect and control your data

Secure data with your encryption key, get granular access control and audit logs, and govern your data with Unity Catalog

### Meet regulatory requirements

A broad set of compliance controls for regulated and sensitive workloads

# Security and compliance






Security and Trust Center - Da

databricks.com/trust

databricks

Platform

Solutions

Learn

Customers

Partners

Company

Try Databricks

Watch Demos

Contact Us

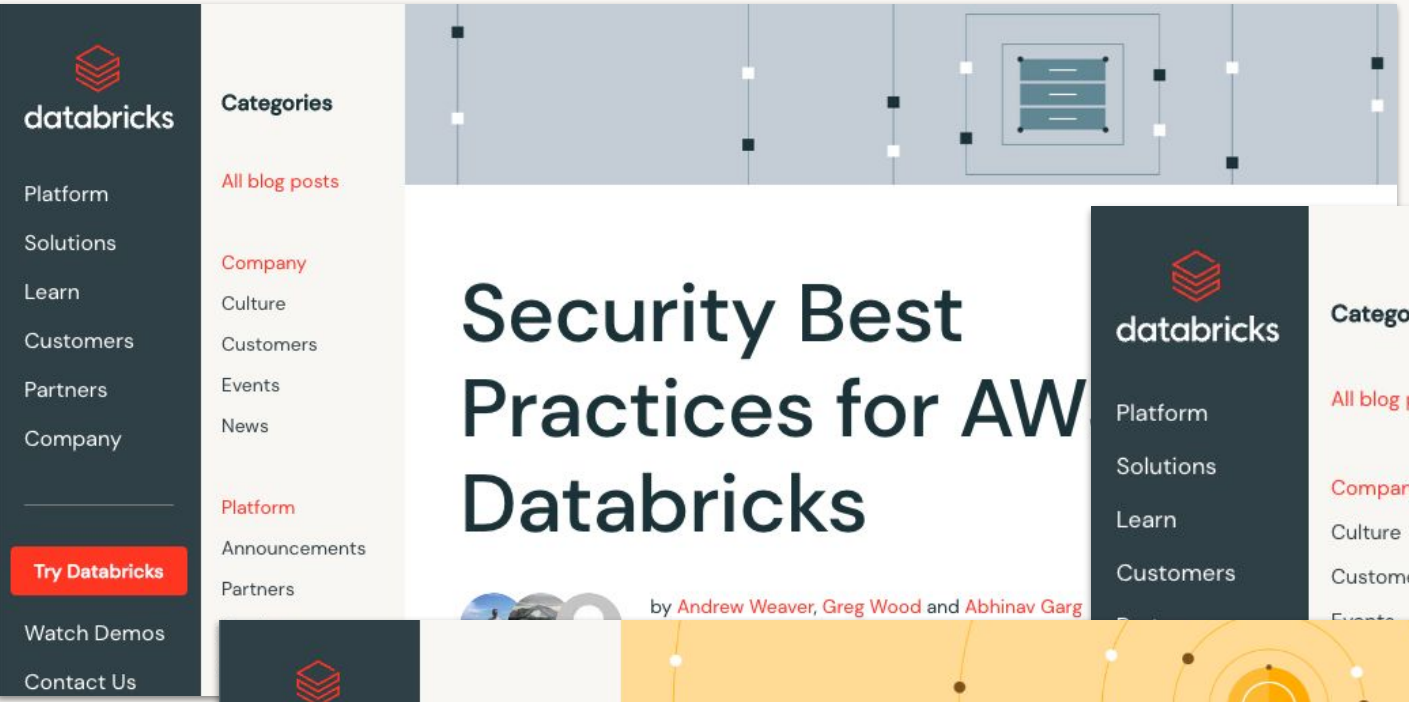
Login

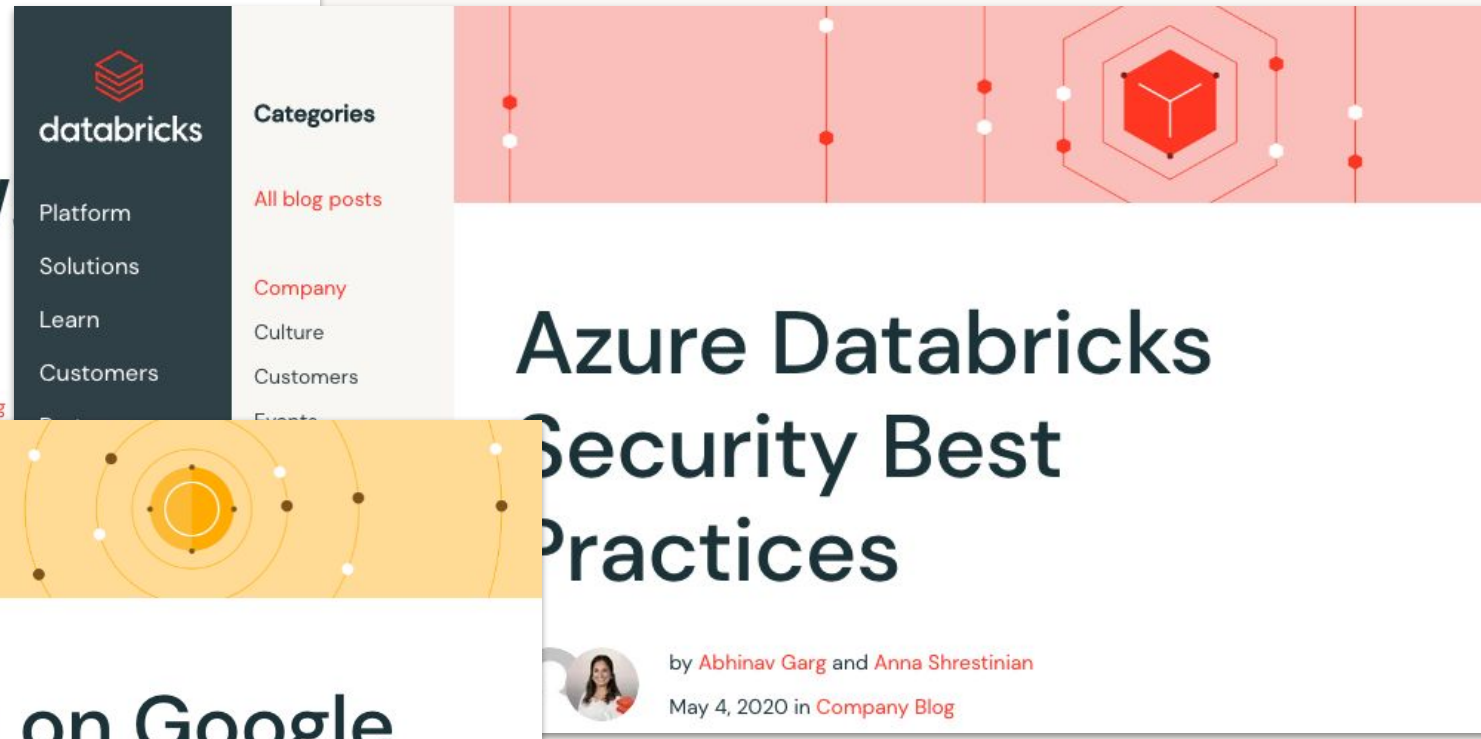
# Security & Trust Center

## Your data security is our priority

Overview | Trust | Security Features | Architecture | Compliance | Privacy

Report an Issue



# Data governance and Unity Catalog

- Unified governance solution build into the Databricks Lakehouse Platform.
- Secure data sharing with Delta Sharing integrated directly into Unity Catalog.
- Single source of truth for data and access controls across all workspaces.
- Provides auditing and data lineage capabilities.
- Existing tables and views can be upgraded to Unity Catalog.

# How Databricks brings down the cost of ownership

# Businesses derive value from Databricks in two ways:

## Value Drivers

1. **Eliminate Cost and Complexity**

2. **Accelerate Innovation**

# We can use these **value drivers** to build a broad, defensible Lakehouse value story

## Value Drivers

**Eliminate Cost and Complexity**

Reduce Total Cost of Ownership (TCO)

**Retire existing software**

- Retire alternative tech with migration to the Lakehouse
- Includes data lakes & warehouses both on prem + in cloud

**Reduce data processing cost**

- Lower cost to process data with Databricks vs. cloud native alternatives
- Proven method with TPC-DS enhanced with POC results

**Data team productivity**

- Reduced cost of data team to both:
  1. **Build**
  2. **Maintain**
- Grounded in Forrester TEI study & customer surveys

**Other cost & complexity value**

- Multi-cloud
- Consolidation with lakehouse
- Improved governance
- … + More to come

"Hard TCO benefits" → Softer TCO benefits

**Accelerate Innovation**

**Value of use cases …**

Streaming, data warehousing, and data science use cases

⬆️ Increase Revenue

⬇️ Decrease Cost

**… Acceleration with Databricks**

How much faster can we unlock this value with Databricks?

Data team productivity → accelerated time to market

# $4.0M of value will be unlocked with Databricks

**Example Scenario**

## Value Summary

Reduce Total Cost of Ownership (TCO)

**Eliminate Cost and Complexity**

### Reduce data processing cost — $0.2M

- $0.1M savings from data engineering pipelines running 66% faster
- $0.1M savings from data analytics running 34% faster

### Increase data team productivity — $1.9M

- 20% improvement in data team productivity across 60 FTEs
- Reduce time to build new data assets and reduce effort to maintain data platform & assets

**$2.1M**

Annual Reduced TCO

**Accelerate Innovation**

### Fraud detection analytics — $0.8M

- Use advanced analytics to detect fraud (eg. credit-card usage)

### Call center call routing — $0.5M

- Route call-center cases based on multimodal data to increase customer satisfaction and reduce handling costs

### Online customer segmentation — $0.5M

- Segment website visitors and present customized pages and offers based on surfing history and customer-specific data

+ Future use cases

**$1.8M**

Annual Accelerated value

**Total annual value**

**$4.0M**

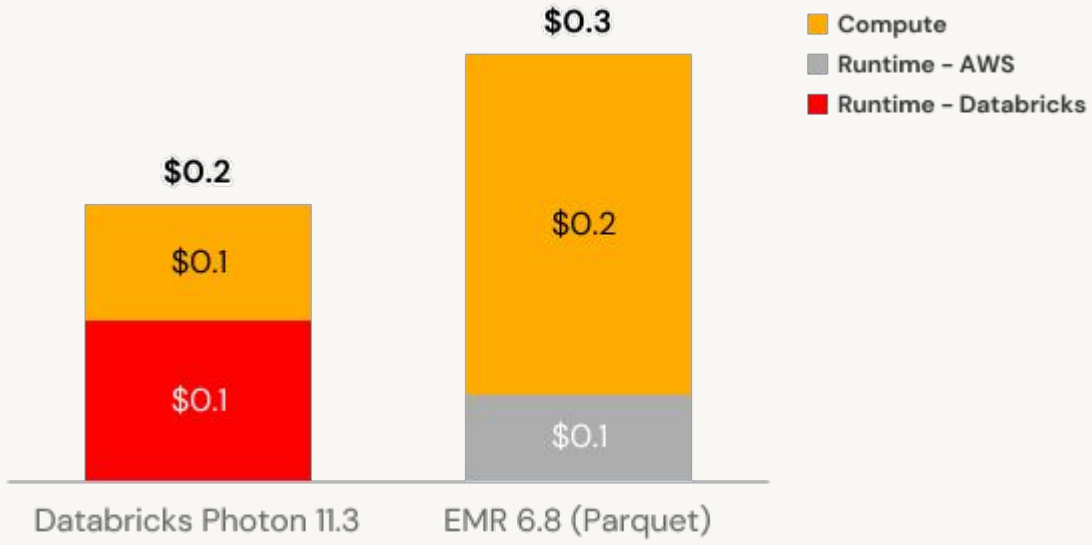
Source: 1. Test Value Forecast Model (2023), 2. Forrester (May 2020), 3. TPC-DS Benchmarks (June 2021), 4. Cloud service pricing (public website)

# Databricks delivers **lower infrastructure costs** and **faster** time-to-insights

**Example Scenario**

## Databricks Photon 11.3 Savings Over EMR 6.8 (Parquet)

$M per year

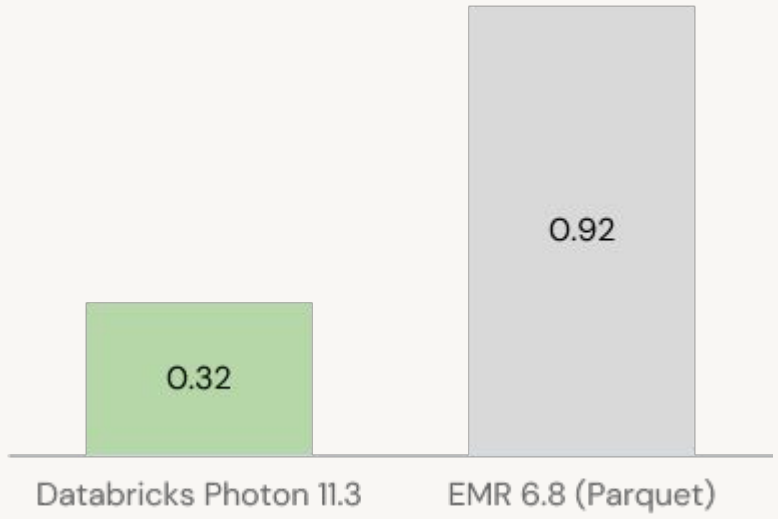



**35%** lower cost

**$0.09K** saving

## Time to insights – 30TB

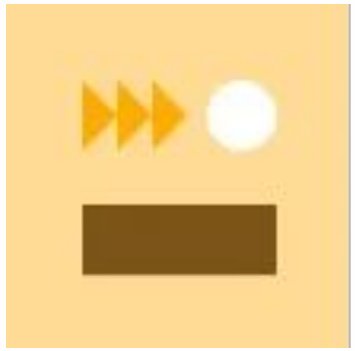
*Hours - Lower is better*




**66% faster**

Source: 1. Test Value Forecast Model (2023), 2. Forrester (May 2020), 3. TPC-DS Benchmarks (June 2021), 4. Cloud service pricing (public website)

# Purchase offerings



Pay as you go

Committed-use

# Product pricing

Workflows & Streaming

## Jobs

Starting at $0.07 / DBU

Run data engineering pipelines to build data lakes and manage data at scale

Learn more →

Workflows & Streaming

## Delta Live Tables

Starting at $0.20 / DBU

Easily build high-quality streaming or batch ETL pipelines using Python or SQL with the DLT edition that is best for your workload

Learn more →

Data Warehousing

## Databricks SQL

Starting at $0.22 / DBU

Run SQL queries for BI reporting, analytics and visualization to get timely insights from data lakes. Available in both Classic and Serverless (managed) Compute.

Learn more →

Data Science & Machine Learning

## All Purpose Compute for Interactive Workloads

Starting at $0.40 / DBU

Run interactive data science and machine learning workloads. Also good for data engineering, BI and data analytics

Learn more →

Data Science & Machine Learning

## Serverless Real-Time Inference

Starting at $0.07 / DBU

Make live predictions in your apps and websites.

Learn more →

Databricks Platform & Add-Ons

## Databricks Platform & Add-Ons

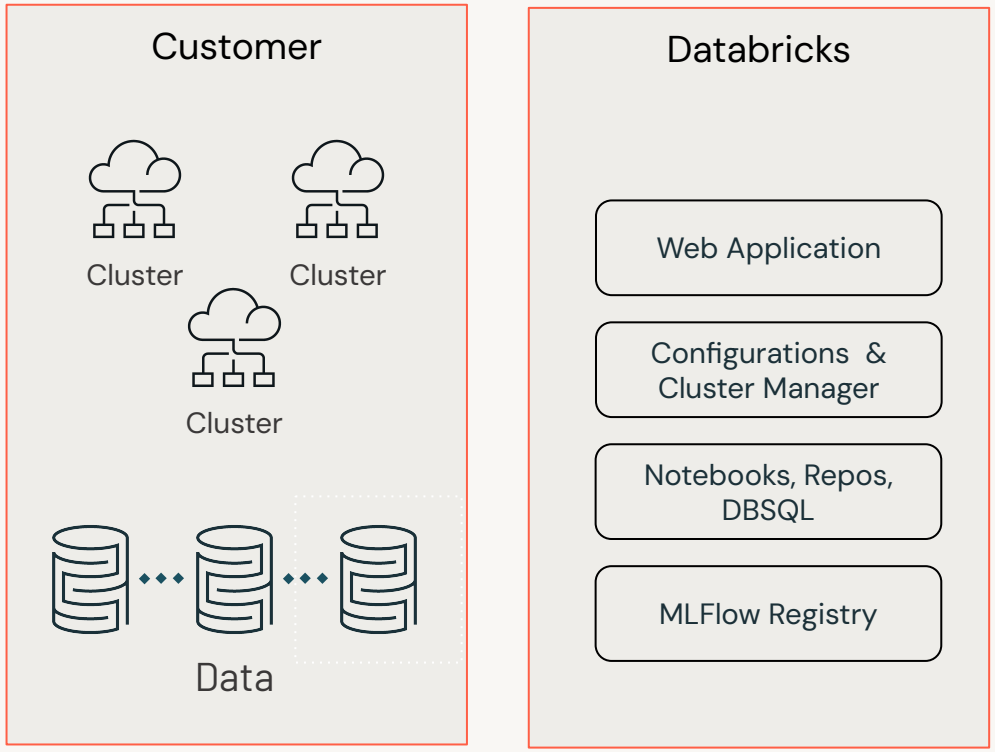
Cross-platform capabilities that provide the right level of management, governance and security to run everything from basic to enterprise-critical workloads

Learn more →

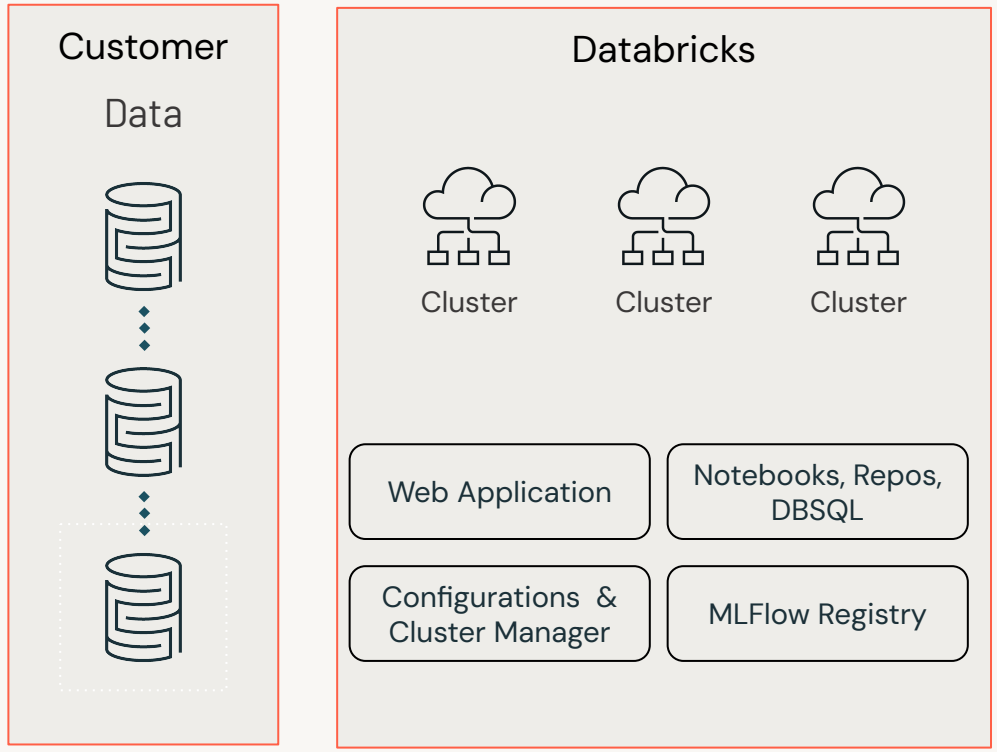
# Compute resources

## For ad-hoc, interactive queries

### Classic compute




### Serverless compute

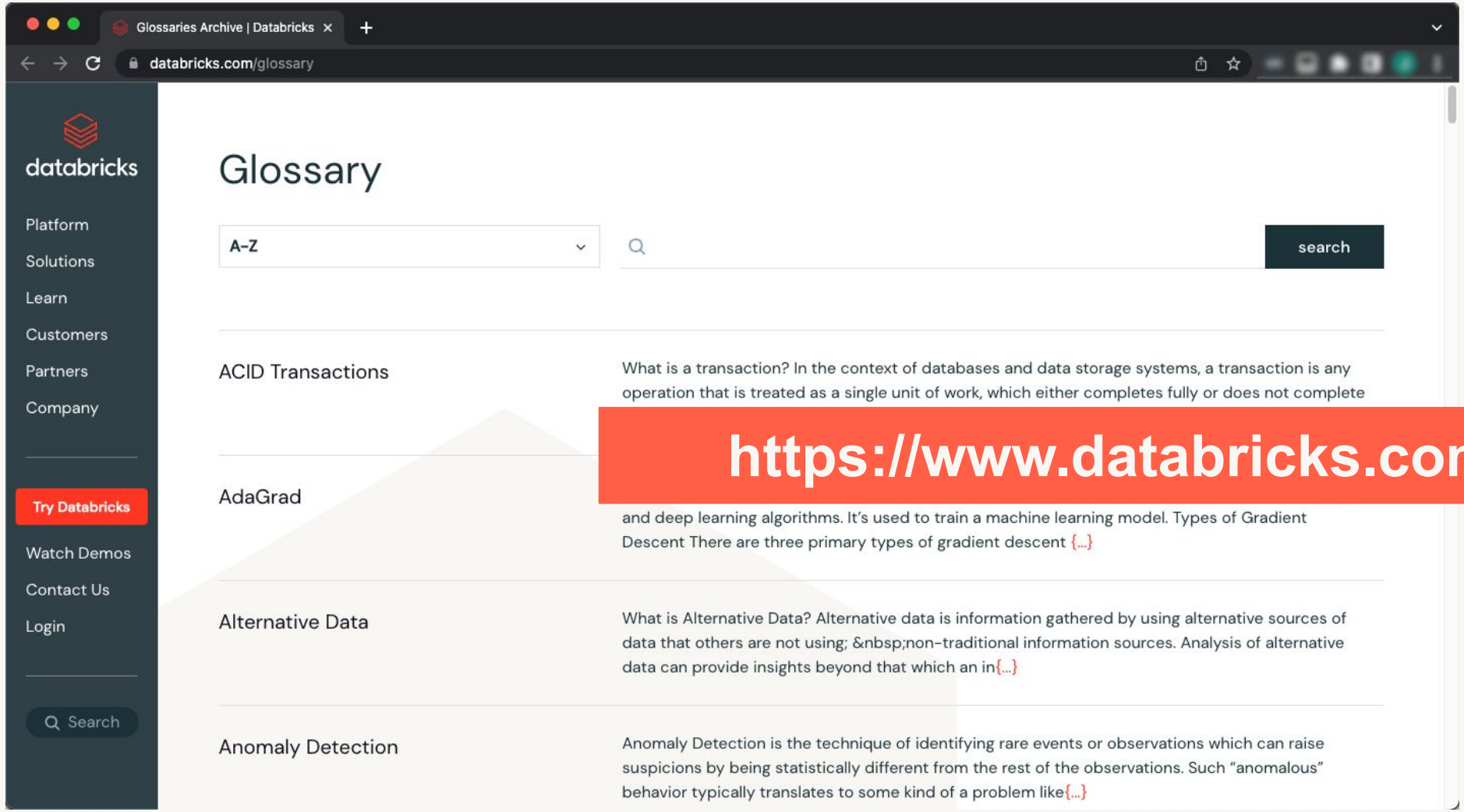


# Summary and next steps

# Why you should include Databricks in your data architecture

- World record holding price performance
- Open standards and formats
- Unified collaborative environment for data teams
- Full support for unstructured data
- Open and secure data sharing
- Enhanced enterprise security and governance
- Simplified production ETL at scale
- Full support for analytics and ML on streaming data
- Open collaboration spaces for all data products

# Databricks Glossary




**https://www.databricks.com/glossary**

# Next steps in your learning journey

- Speak with your Databricks Account Executive to learn more about how you can implement Databricks with your unique requirements.
- Visit Databricks Academy to view additional courses.
- Join the Databricks Community to connect with other Databricks customers.
- Check your email for future Databricks Workshops like this one in the future!

# Thank you!